DOCUMENT RESUME

ED 417 726 IR 056 951

AUTHOR Phillips, Lori
TITLE Comparison of the Knight-Ridder CD-ROM and CARL Versions of the ERIC Database.
PUB DATE 1997-05-00
NOTE 14p.; Paper presented at the Oklahoma Library Association and the Mountain Plains Library Association Joint Conference (Afton, OK, April 30-May 3, 1997).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Libraries; Access to Information; Computer System Design; *Databases; Decision Making; Higher Education; *Information Retrieval; *Library Automation; Library Collection Development; *Online Searching
IDENTIFIERS *ERIC

ABSTRACT

This study, conducted at the University of Wyoming (UW), illustrates the complexity involved in making reasonable decisions about which electronic products to make available in the library. UW is a member of the CARL Alliance, and utilizes the CARL integrated online system which includes an OPAC, circulation, serials control, bibliographic maintenance and acquisitions modules. When the CARL Alliance acquired and loaded ERIC in April of 1992, UW Libraries added it to the OPAC menu. Striking differences were noticed in the amount of information retrieved by the two systems. The concept of descriptors is key to understanding how the ERIC database is structured. Both keyword and controlled vocabulary searches are accommodated by Dialog (CD-ROM). For the CARL interface, a Word search retrieves words from all fields of a record; while Subject Word and Subject Browse searches retrieve major descriptors only. This study was undertaken with the idea that some kind of systematic and orderly collection of data was needed in order that librarians might make an informed comparison of the capabilities of the two systems. User education relating to ERIC on CARL is called for on the part of libraries finding themselves with no other access to the database; the incomplete sets of retrieved information are cause for concern. Users need to be aware of the shortcomings of CARL, and institutions with large numbers of off-campus students have a particular responsibility to inform their clients of potential problems with their searches. Librarians should evaluate their own working environments to determine if impediments to access such as those found at University of Wyoming exist in their local situations. (Contains 21 references.) (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 417 726

Comparison of the Knight-Ridder CD-Rom and CARL Versions of the ERIC Database

By Lori Phillips
University of Wyoming Libraries

Paper presented at the Oklahoma Library Association and the Mountain Plains Library Association Joint Conference

Academic Research Forum
Sponsored by the MPLA Academic Section

Shangri-La Resort
Grand Lake O' the Ozarks
Afton, Oklahoma
April 30-May 3, 1997

IR056951

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Lori Phillips
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

CONTEXT

One professional responsibility of librarians is to evaluate tools to determine if and how they will be useful to their clients. As more resources become available electronically, the discerning professional eye that has focused on print resources examines electronic resources as well. Decisions about resource provision move from the realm of "acquire or don't acquire" to acquire in which format, and from which vendor. This study, conducted at the University of Wyoming, illustrates the complexity involved in making reasonable decisions about which electronic products to make available.

According to the last census, Wyoming in the most rural state in the union. There is only one four-year institution of higher education in the state, the University of Wyoming. It serves a student population of approximately 10,000. Of these, 8,000 are undergraduates and 2,000 are graduates. Many of these 10,000 students acquire their education at a distance, as 2,325 students are enrolled in off campus courses for the Spring 1997 semester.

Library resources to serve this student population are varied. The University of Wyoming Libraries consists of one main library and four branches. The collection contains over one million volumes and nine thousand journal subscriptions. UW is a member of the CARL Alliance, and utilizes the CARL integrated online system which includes an OPAC, circulation, serials control, bibliographic maintenance and acquisitions modules.

At the time data was collected for this study over 100 databases, including library catalogs, article indexes, information databases and gateways to other library systems were available to UW users through CARL. The library also provided access to over seventy cd-rom

titles, some commercially produced, and others produced by the federal government. These cd-roms were not networked; multiple titles ran on stand-alone workstations.

ERIC has always been one of the most heavily used databases at University of Wyoming. Sign-up sheets and long delays to access the database at its' stand-alone workstation were the rule rather than the exception. When the CARL Alliance, on behalf of its members, acquired and loaded ERIC in April of 1992, the University of Wyoming Libraries did not hesitate to add it to the OPAC menu. It was an attractive option for two reasons. First, it allowed us to move from a single, library located workstation to simultaneous, distributed access. Given the previous congestion problem, this was seen as a major advantage. Another anticipated advantage of accessing ERIC via CARL was that users would be able to rely upon what they already knew about CARL search protocols when using ERIC. Other libraries that had loaded commercial databases into their home systems reported that these factors were considerations for them. For example, when Indiana State University in Terre Haute decided to mount ERIC in 1990, they chose a NOTIS interface, since they had been a NOTIS library since 1985. NOTIS was willing to work with ISU toward developing a loader program for the ERIC database, and they were able to maintain a common interface across their catalog. CARL likewise developed a program to load ERIC. Patricia Culkin, then Associate Director of CARL, spoke in 1989 about the process whereby an institution can site license a database, load it into their local database system, and apply their local indexing and access software to allow users to search these files parallel with the local collection.

> "Technical issues usually relate to the data itself. While the structures are often MARC or MARC-like, they tend to differ from the standard formats at the field and subfield level, so software work needs to be performed to make them load and index correctly into the local database" (Culkin, p. 175).

Results of the current study indicate that although she correctly identified the challenge, CARL software engineers were unable to develop a product that could respond effectively to that challenge.

Local response to having ERIC available on the OPAC was enthusiastic. Faculty had access from their offices, and students no longer had to sign up for a specific time slot to use ERIC. However, questions about the new system began to surface almost immediately. Striking differences were noticed in the amount of information retrieved by the two systems. There were many anecdotal exchanges between librarians and users about unexpected results of ERIC searches performed using CARL. Some users assumed that the databases were different, since they would find an item using cd-rom and not locate it using CARL.

At the same time, shrinking monetary resources caused members of collection development management to question the necessity for paying for two electronic versions of the ERIC database. These persons argued that CARL access to ERIC should be sufficient. The value of Dialog's sophisticated search software diminished when compared to the ability of CARL to support simultaneous and remote use.

It was against this backdrop that the present study was undertaken. Some kind of systematic and orderly collection of data was needed in order that librarians might make an informed comparison of the capabilities of the two systems. An applied research project was undertaken to gather needed information.

LITERATURE REVIEW

Nash (1990), Kirby (1986), and McCarty (1997) have investigated end user satisfaction with cd-roms. Nash reports user satisfaction with search results even when few useful citations are retrieved. Kirby observes users feel satisfied without understanding that their search results are flawed in terms of quantity and quality.

Many studies have been conducted about various aspects of the ERIC database. Barbuto (1991), Bevis (1996), Hsieh-Yee (1993), Lancaster (1994), Nissley (1989), Puttapithakporn (1990), and Spragg (1992) discuss end user searching and user satisfaction. User satisfaction with ERIC is uniformly high, despite mixed results in terms of both quantity and quality of items found. ERIC Thesaurus structure and use was investigated by Nicholls (1989), and Harter (1996). Reese investigated various software versions of the ERIC database and attempted to give librarians some concise information to use as they made collection development decisions. She concluded in 1987 that Dialog was both "accurate and comprehensive", and noted in 1989 that improvements made it a "first rate" product.

Culkin (1989), Davis (1991), and Fiscella (1995) describe the experience of loading the ERIC database into local catalogs. Fiscella's study indicated that faculty had neither a clear recollection of the two different versions of ERIC loaded into the local system or a clear preference for one version over the other.

METHODOLOGY

To investigate the perception that the two electronic versions of ERIC were somehow "different", a random list of 352 main-entry descriptors was generated using the **Thesaurus of**

ERIC Descriptors. The controlled vocabulary system that is the backbone of the ERIC database, the 1995 edition of the Thesaurus contains 10,363 terms, 5,759 of which are main entry descriptor terms and 4,604 of which are dead terms and "use" terms. From the random list of terms, descriptors were searched two ways, individually and as pairs. Every attempt was made to assure uniformity. The number of records retrieved by the cd-rom and by CARL was recorded for each search. Results were compared using descriptive statistical techniques.

The concept of descriptors is key to understanding how the ERIC database is structured. Each item entered into the database is assigned any number of major and minor descriptor terms. Major descriptors denote a primary subject relationship, while minor descriptor status indicates a subject relationship that that is not as direct. Minor descriptors do however, indicate a subject relationship that is stronger than a key word match. The process of distinguishing between major and minor terms is known as "weighting".

The two systems deal with the structure of the database differently. Dialog's menu driven software allows users to chose either a Word Phrase search or an ERIC Subject Heading search. System documentation explains that the Word Phrase search finds records with the chosen term in any field of the record, while the Subject Heading search returns records with the chosen term found in either the major or minor descriptor field. Both keyword and controlled vocabulary searches are thus accommodated.

For the CARL interface, Word, Subject Word and Subject Browse searches are available. According to CARL's PAC documentation, searches work in the following ways: a Word search retrieves words from all fields of a record; while Subject Word and Subject Browse searches retrieve major descriptors only.

Comparability of the commands on each system can now be established. The Word Phrase command on cd-rom searches the major and minor descriptor fields as well as all other fields of the record. This can be compared to the Word command on CARL, which should read all fields of the record.

The Subject Heading command on cd-rom reads the major and minor descriptor fields. The Subject Word and Subject Browse commands on CARL read the major descriptor field only. Thus, even before formal investigation began, it was obvious on the basis of system documentation that the number of articles retrieved would be skewed in favor of the cd-rom.

DATA

To test the capabilities of each system, the first set of comparisons was made on individual descriptors. Each system provides two options for searching terms, descriptor and keyword. For descriptor terms, the Subject Heading (cd-rom) and Subject Word (CARL) searches were used. Of the 352 terms searched, the cd-rom found more records per search in every instance. This came as no surprise, since CARL only looks for major descriptors.

Next, keyword searching of single terms was investigated. The Word Phrase (cd-rom) and Word (CARL) searches were used. The expectation was that the results would be comparable, based on system documentation. This did not, however, prove to be the case. Of the 352 single terms searched, the cd-rom found more records per search for 282 of those terms, or 80% of the total. This result was puzzling and warranted further investigation.

Sets that contained a relatively small number of records were chosen for closer scrutiny. Comparison of individual records retrieved by the two systems provided an answer. The word

search on CARL, which supposedly accessed every field in the record, systematically left off the minor descriptor field from consideration. Operationally, the word search was looking at the major descriptors, words in titles, abstracts and elsewhere in the record, but left off minor descriptors. Indeed, in the subsample of individual records that was examined, the ones that were missed by CARL all had the term in the minor descriptor field.

Taking the analysis one step further, the number of records found by cd-rom for each individual term was used as a baseline number. The number of records retrieved by CARL was then computed as a percentage of possible relevant records. When these individual percentages were averaged for the 352 terms, using both keyword and descriptor samples, it was discovered CARL found only 56% of the possible relevant records.

Many end users of the ERIC database attempt to combine two terms in a single search. To test system capabilities in this area, pairs of descriptor terms were constructed from the random list of terms. While performing the paired term searches, it was discovered that the order in which the terms were entered into CARL influenced the search results. To illustrate, entering the terms "incentives" and "productivity" in one order (a+b) found fifteen records using CARL; searching the terms in the opposite order (b+a) found twenty one records. The combinations of "incentives" and "tests" found nine records in one order (a+b) and no records in the opposite order (b+a).

Results of the paired term searches conform to the previously established pattern. Of 336 pairs searched, the cd-rom found more information in 75 % of searches, CARL in 21% of searches, with the number of records being equal in 4% of searches.

DISCUSSION

Studies indicate users may forego use of controlled vocabulary in favor of natural language, and do not distinguish between keyword and descriptor searching. However, controlled vocabulary does allow us to test the capabilities of search software, by providing as a benchmark all records indexed using a particular term against which retrieval can be measured. Dialog takes full advantage of the controlled vocabulary structure of ERIC. It allows for searching by both major and minor descriptors, as well as a keyword search. It is a reasonable standard against which we can measure the performance of CARL.

The approach that CARL has taken to the ERIC database is best understood by exploring the general corporate approach to information retrieval. CARL did not begin as a corporation but rather as a group of library directors interested in interlibrary cooperation. Formed in 1974, early efforts included a union list of serials, and cooperative purchasing of expensive materials. It incorporated as a private not for profit corporation in 1978, and began the metamorphosis from consortium into corporation. Ward Shaw, Executive Director of CARL, Inc. articulated the corporate philosophy concerning information retrieval in 1980 when he wrote,

> "In the case of systems for public access...success will not be measured by such easily quantifiable things as relevance/recall ratios or numbers of documents delivered, but rather by user satisfaction," (Shaw, p3).

The editorial decision made by CARL to eliminate the minor descriptor field from consideration when implementing their version of the database has demonstrable consequences. Patrons are unable to locate all information relevant to their queries. Successful searching for more than one term in a single search statement depends upon the order in which the terms are

entered. CARL's approach to information retrieval does not take advantage of the structure of the ERIC database.

The decision about what level of deficient performance will be tolerated in online products is one that each individual library must make. Using the study as one strand of information in the decision making process, it was decided to retain both electronic versions of the ERIC database for the short term, the one for its sophisticated search software, and the other for its availability to multiple users, some of which were located outside the library building. However, the long-term goal was to find a way of providing distributed access to the ERIC database other than CARL. In January of 1997, the cd-rom was replaced with SilverPlatter's ERL version of ERIC, which was mounted on the OPAC. Despite the fact the University of Wyoming declined in 1996 to continue to pay for the ERIC database as part of our agreement with CARL, it is still available, by virtue of a CARL Alliance decision to maintain it as part of a basic package. We currently provide no direct pointer to what we jokingly refer to as the "database that will not go away".

Both reference work and user education at the University of Wyoming were impacted by this study. Librarians gained confirmation that ERIC on CARL provided users with retrieved sets that were different than the sets retrieved using the cd-rom, using indentical search strategies. Graduate students and faculty in the College of Education were targeted and made aware of the limitations of ERIC on CARL; the former through instructional sessions connected with their introductory research methods class and the latter through faculty meetings. Outreach librarians faced a particular challenge in dealing with distance students that had access only to the CARL version of ERIC.

IMPLICATIONS

User education relating to ERIC on CARL is called for on the part of libraries finding themselves with no other access to the database. The incomplete sets of retrieved information are cause for concern. Users need to be aware of the shortcomings of CARL. Institutions with large numbers of off-campus students have a particular responsibility to inform their clients of potential problems with their searches.

Librarians should evaluate their own working environments to determine if impediments to access such as those found at University of Wyoming exist in their local situations.

BIBLIOGRAPHY

Barbuto, Domenica M. and Elena E. Cevallos. "End-user searching: program review and future prospects", **RQ**, v. 31, no. 2: 214-227 (Winter 1991).

Bevis, Mary D. "ERIC on CD-ROM and RIE: A comparison of search results and user satisfaction", **The Reference Librarian**, no. 53: 105-112 (June 1996).

Culkin, Patricia B. "Rethinking OPACS: the design of assertive information systems", Information **Technology and Libraries**, v. 8, no. 2: 172-177 (June 1989).

Culkin, Patricia B., and Ward Shaw. "The CARL (Colorado Alliance of Research Libraries) system", Library **Journal**, v. 110, no. 2: 68-70 (February 1, 1985).

Davis, H. Scott and Pat L. Ensor. "Making the transition to the SuperCatalog: ISU's experience with LUIS II", College **& Research Libraries News**, v. 52, no.2: 94-96 (February 1991).

Fiscella, Joan B., and Edward Proctor. "An approach to assessing faculty use of locally loaded databases", **College & Research Libraries**, v. 56, no. 5: 446-458 (September 1995).

Harter, Stephen P., and Yung-Rang Cheng. "Colinked descriptors: improving vocabulary selection for end-user searching", **Journal of the American Society for Information Science**, 47, (April 1996): 311-25.

Hsieh-Yee, Ingrid. "Effects of search experience and subject knowledge on the search tactics of novice and experience searchers", **Journal of the American Society for Information Science**, v. 44, no. 3: 161-174 (April 1993).

Kirby, Martha and Naomi Miller. "MEDLINE searching on Colleague: reasons for failure or success of untrained end users", Medical **Reference Services Quarterly**, v. 5, no. 3: 17-30 (Fall 1986).

Lancaster, F. Wilfrid, Cheryl Asper Elzy, and Mary Jo Zeter. "Searching databases on CD-ROM: comparison of the results of end-user searching with results from two modes of searching by skilled intermediaries", **RQ**, v. 33, no. 3: 370-386 (Spring 1994).

Lenzini, Rebecca T., and Ward Shaw. "Creating a new definition of library cooperation: past, present, and future models", Library **Administration & Management**, v. 5, no. 1:37-40 (Winter 1991).

McCarty, Cheryl A., Sylvia C. Krause, and Arthur A. Little. "Expectations and effectiveness using cd roms: what do patrons want and how satisfied are they?", College **& Research Libraries**, v. 58, no .2: 128-142 (March 1997).

Nash, Stan and Myoung Chung Wilson. "Value-added bibliographic instruction: teaching students to find the right citations", Reference **Services Review**, v. 19, no.1: 87-92 (Spring 1991).

Nicholls, Paul and Susanne Holtmann. "Women's issues searching with DIALOG OnDisc ERIC: natural language and controlled vocabulary strategies", The **Laserdisk Professional**, v. 2, no. 3: 97-103 (May 1989).

Nissley, Meta, Peter G. Anderson and Phyllis Gaal. "ROMping through ERIC: measuring satisfaction and effectiveness", **The Laserdisk Professional**, v. 2, no. 1: 95-100 (January 1989).

Puttapithakporn, Somporn. "Interface design and user problems and errors: a case study of novice searchers", **RQ**, v. 30, no. .2:195-204 (Winter 1990).

Reese, Jean. "ERIC on compact disk: new software versions from Dialog, OCLC & SilverPlatter", **The Laserdisk Professional**, v.2, no.1: 75-80 (January 1989).

Reese, Jean and Ramona Steffey Madewell. "ERIC on CDROM: a comparison of DIALOG OnDisc, OCLC's Search CD450 and SilverPlatter", **Online**, v.11, no. 5: 42-54 (September 1987).

Rice-Jones, Judith. "An interview with Card Shaw and Patricia Culkin: designers of CARL Pac", Colorado **Libraries**, v. 11: 5-12 (March 1985).

Shaw, Ward. "Design principles for public access", In **Public Access to Library Automation, Clinic on Library Applications of Data Processing**, 17th edition, Urbana, Ill.: University of Illinois, 2-7,1980.

Spragg, Edwin B. "CD-ROM software evaluation; comparison of searching ERIC on IBM and Macintosh", **College & Research Libraries**, v 53, no 4: 342-351 (July 1992).